AD-A042 789 IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS N Y F/G 12/1
ILL-POSED PROBLEMS, REGULARIZATION AND SINGULAR VALUE DECOMPOSI--ETC(U)
APR 77 J CULLUM F44620-76-C-0022
UNCLASSIFIED RC-6465 AFOSR-TR-77-0880 NL

1 OF 1
ADA042 789

END
DATE FILMED
9-77
DDC

(14) RC-6465 (#27919) 4/5/77
Mathematics 37 pages

(17) AFOSR-TR- 77-0850

(11) AFOSR

Research Report

ILL-POSED PROBLEMS, REGULARIZATION AND SINGULAR VALUE DECOMPOSITIONS

Jane Cullum

(11) 5 Apr 77

(9) Interim rept.,

(12) 41 p.

(15) F44620-76-C-0022

(16) 2304 (17) A4

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Typed by Marilyn McCrae on CMS (jc.1018)

IBM Research Division
San Jose · Yorktown · Zurich

RC 6465 (#27919) 4/5/77
Mathematics 37 pages

ILL-POSED PROBLEMS, REGULARIZATION AND SINGULAR VALUE DECOMPOSITIONS

Jane Cullum

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Typed by Marilyn McCrae on CMS (jc.1018)

Abstract: We consider ill-posed problems of the form $g(t) = \int_0^1 K(t,s)f(s)ds$ and their discrete approximations obtained by quadrature, $Ax = b$. We assume that our desired solution f is smooth and that our data g is measured experimentally and contains highly oscillatory noise. With theorems and examples we demonstrate the effect of each of these procedures, the singular value decomposition with truncation, (SVDT) a Hankel transformation with damping, and the Tikhonov regularization procedure, on such noise in the data. We demonstrate that in general, regularization is the most natural setting for mollifying the effects of such noise. However, for certain problems SVDT is equally suitable and in fact may be better if the rate of convergence of the regularization procedure is too slow.

* This Research was sponsored by the Air Force Office of Scientific Research (AFSC), United States Air Force, under contract F44620-76-C-0022.

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication elsewhere and has been issued as a Research Report for early dissemination of its contents. As a courtesy to the intended publisher, it should not be widely distributed until after the date of outside publication.

Copies may be requested from:
IBM Thomas J. Watson Research Center
Post Office Box 218
Yorktown Heights, New York 10598

1. Introduction: We consider the following type of ill-posed problems. In the continous case we consider a linear integral equation of the first kind

$$g(t) = \int_0^1 K(t,s)\, f(s)ds \tag{1}$$

where K is an L_2- kernel and we assume that the null space of K, N(K), is empty. Then K is a compact operator, and its inverse is unbounded. In the algebraic case we consider a square $n \times n$ system of linear equations

$$Ax = b \tag{2}$$

where A is invertible, but ill-conditioned. As a measure of ill-conditioning we use the ratio of the largest to the smallest singular values. In our discussions we will consider (2) as a discrete approximation to (1), obtained using a quadrature rule. Some papers, for example, Faddeev and Faddeeva [1], Zhukovskii and Morozov [2], and Tikhonov [3] consider the regularization of general systems (2) not necessarily related to a continuous problem (1). However, we will see that in the context of the current discussion it may not be reasonable to consider general systems.

A problem is ill-posed if small changes in the data - in problem (1) the data is g(t), in problem 2 it is b - can yield large changes in the solution.

Example. In (1) let $K(t,s) = \begin{matrix} 0 & t < s \\ 1 & t \geq s \end{matrix}$.

Then solving (1) is equivalent to differentiating g. If we replace g by $g_\epsilon = g + \epsilon \sin\omega t$ for small ϵ and large ω, we see that for the solution of (1), f_ϵ corresponding to g_ϵ, the error $f_\epsilon - g' = \epsilon\omega \cos\omega t$, is large.

The application of standard numerical analytic techniques to the solution of (1) or (2) yields non physical, highly oscillatory solutions. Equations of the form (1), arise in many applications, for example the numerical differentiation of tabulated data, the deconvolution of data obtained in spectroscopy experiments, inverse problems in geophysics, and in signal processing, such as radar and sonar. For examples of some of these applications see the following references Bachmann et al [4], Huang and Parrish [5] and Backus and Gilbert [6].

We emphasize the fact that we want a solution, not just at one point of an interval, but over the entire interval. Moreover, we are dealing with experimentally determined data so this data almost surely contains errors, perhaps both systematic and random. The following discussion assumes that any systematic errors have been removed from the data supplied to (1). In many cases, the random error obtained is highly oscillatory and of low amplitude, and the desired measurement is smooth. It is this situation that we want to discuss.

In recent years several schemes, see for example, Hanson [7], Varah [8], Ekstrom and Rhoads [9], Tikhonov [10], and Lee et al [11] have been proposed for solving (1) and (2). We will discuss 3 of these, the singular value decomposition with truncation (SVDT) Hanson [7], Varah [8]; the Hankel transformation procedure of Ekstrom and Rhoads [9]; and the regularization procedure of Tikhonov [10]. We will attempt to give some understanding of the relationships between these 3 procedures and of the strengths and weaknesses of each of these procedures by presenting a few Theorems and examples. We will argue, using ideas from Anderssen and Bloomfield [12] for numerical differentiation, that of the 3 procedures, the regularization procedure yields the most natural resolution of the problem of handling noisy data. We argue by example that in general, the SVDT procedure may not mollify the effects of the noise in the data. The Ekstrom-Rhoads procedure lies somewhere between SVDT and regularization. This Hankel transformation procedure is very interesting and does mollify the noise. We can, however, show by example that it may also mollify the desired solution which is something we do not want to do. The key to our problem is an appropriate choice of basis,

see for example Varah [13]. To decide what basis is appropriate we must first decide what the objective of our computation is. What would we like to be able to do?

First we state two basic premises.

Premises

1) We assume that the desired solution is smooth with one or more derivatives. 2) We assume the data contains highly oscillatory additive noise. See Figure 1.

Given these premises our objective is as follows.

Objective

To approximate our original problem by a better conditioned problem that (1) reduces the influence of the noise, and (2) gives us a physically meaningful solution that approximates the true solution in some reasonable sense. We will demonstrate by example in the section on SVDT that modifying a problem so that it is numerically stable does not necessarily guarantee that we have simultaneously mollified the significant part of the noise. Clearly, the errors in the data impose a limitation on the achievable accuracy.

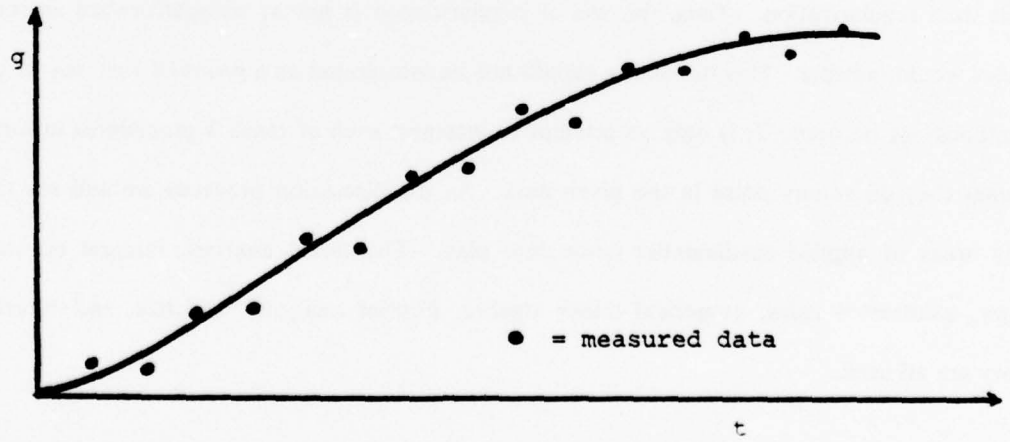


Figure 1

We note in passing that in certain situations, for example in the registration system of an electron beam column Wilson et al [14] noise removal is achieved by simply running the same experiment many times and then averaging the data obtained over the many runs. With truly

random data with zero mean averaging over a set of experiments will eliminate the noise. However, there are many situations where it is impractical or even wrong to run the same experiment many times. See Chang [15] for example which discusses Auger spectroscopy. In Auger analysis a beam of electrons is shot at a specimen. Certain types of electrons subsequently emitted from the specimen are collected and used to determine the chemical composition of the specimen. However, each time the specimen is exposed to the beam, deterioration of the specimen occurs and in fact the sample at the end of an experiment may not be identical to what it was in the beginning. Therefore, noise averaging is not feasible.

As stated earlier, in the preceeding context we will discuss 3 methods SVDT, Ekstrom and Rhoads, and regularization keeping the 2 premises and our stated objectives in mind. We conclude that regularization is the most direct of the 3 methods for handling any random noise in the data. We note, however, that in certain situations SVDT also effectively handles the noise. We further note that the goodness of a regularization approximation depends strongly upon the rate of convergence of the approximations, and for some problems this rate as pointed out by Franklin [16] can be very slow. In such a situation a SVDT may yield a better result than regularization. Thus, the use of regularization is not as straightforward as some articles would indicate. This discussion should not be interpreted as a proposal that any of the 3 methods not be used. It is only an attempt to interpret each of these 3 procedures in terms of what they do to any noise in the given data. As the discussion proceeds we will see that many areas of applied mathematics come into play. Functional analysis, integral equation theory, quadrature rules, numerical linear algebra, Fourier analysis, statistics, and filtering theory are all used.

2. The Singular Value Decomposition with Truncation (SVDT).

Picard's Theorem, Smithies [17], gives necessary and sufficient conditions for the existence of a solution to (1).

Theorem 1 [17]. Let $K(t,s)$ be an L_2 - kernel for (t,s) in $[0,1] \times [0,1]$, and let $g \in L_2[0,1]$. Then (1) has an L_2 - solution f if and only if

$$\sum_{n=1}^{\infty} \frac{1}{\sigma_n^2} (g^T u_n)^2 < \infty \tag{3}$$

and g is in the closure of the range of K, $\overline{R}(K)$. In (3) u_n, is a full set of eigenfunctions for the operator $K K^*$ and σ_n^2 is the corresponding set of eigenvalues, $\sigma_n^2 \neq 0$ as $n \to \infty$. K^* is the adjoint operator of K, and its kernel is $K^*(t,s) = \overline{K(s,t)}$.

Theorem 2 [17]. Under the hypotheses of Theorem 1, equation (1) has the unique solution in the orthogonal complement of $N(K)$,

$$f = \sum_{n=1}^{\infty} \frac{1}{\sigma_n} (g^T u_n) v_n \tag{4}$$

where the v_n are the eigenfunctions of the operator K^*K.

The system $\{u_n\}$, $\{v_n\}$, $\{\sigma_n\}$, $n = 1,2,\ldots$ is the singular value decomposition of the operator K, Smithies [17].

That is,

$$K = \sum_{n=1}^{\infty} \sigma_n u_n v_n^T \tag{5}$$

Similarly, for the matrix A, we obtain

$$A = U \Sigma V^T \tag{6}$$

where the columns of U and V are the vectors $u_1,\ldots,u_n$ and $v_1,\ldots,v_n$ respectively, and Σ is the diagonal matrix whose non zero entries are σ_j, $1 \leq j \leq n$. The σ_j are called the singular values of K or A. We note that the closure of the span of the v_n , $n = 1,2,\ldots$ is the closure of the range of K^* , $\overline{R(K^*)}$, which equals the orthogonal complement of the null space of K,

$N(K)^{\perp}$. Moreover, the closure of the span of the u_n, $n = 1,2,...$ equals the closure of the range of K, $\overline{R(K)}$, which equals the orthogonal complement of the null space of K^*, $N(K^*)^{\perp}$. Since we assume $N(K) = \{0\}$, we have that the v_n span all of L_2.

We note that there is a generalization of the SVD due to Van Loan [18] for real matrices of the form A + B. It is called the BSVD and uses 3 matrices U,X, and V with U and V orthogonal matrices and X nonsingular, plus 2 diagonal matrices Σ_A and Σ_B. The resulting decomposition is

$$\begin{aligned} U^T A X &= \Sigma_A \\ V^T B X &= \Sigma_B \end{aligned} \tag{7}$$

If we denote the non zero entries in Σ_A by α_i $1 \le i \le n$, and those in Σ_B by β_i $1 \le i \le n$, then the B - singular values of A

$$\mu(A,B) = \{\mu \mid \mu \ge 0, \det (A^T A - \mu^2 B^T B) = 0\}$$

are given by

$$\mu(A,B) = \{\alpha_i/\beta_i \mid 1 \le i \le n\} \ .$$

We have stated both the SVD and BSVD results for square matrices, but both apply to $m \times n$ rectangular matrices. The BSVD with truncation does not seem to have been used in the literature, one exception is Varah [13]. We will not directly consider the BSVD; however, the comments that we make about the SVD with truncation also apply to a BSVD with truncation. The basis of interest in that case is the columns of X.

The SVD with truncation (SVDT) is easier to discuss in the algebraic framework (2), so we consider (2). For any real matrix A we have a SVD

$$A = U \Sigma V^T \tag{8}$$

where U and V are orthogonal matrices, U an ortho-normal set of eigenvectors of AA^T, V an ortho-normal set of eigenvectors of A^TA and Σ a diagonal matrix whose entries are the square roots of the eigenvalues of AA^T, $\sigma_j \downarrow$ as $j \uparrow n$. The columns in U and V are ordered to match the ordering in Σ. In our problems A is square and nonsingular, so none of the diagonal entries in Σ vanish.

A SVDT is achieved by considering the entries in Σ and deciding which entries are 'not significant' Varah [13], Lawson and Hanson [19]. 'Insignificant' entries are set equal to zero

obtaining $\overline{\Sigma} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$ from

$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$. Then the equation

$$\overline{A}x = U\overline{\Sigma}V^Tx = b \tag{9}$$

is solved.

This approach assumes that the projection of the data onto the singular vectors corresponding to those singular values reduced to zero is small. In fact it assumes more. Namely, that the projection g^Tu_n amplified by $1/\sigma_n$ is small. Varah [9] discussed this in detail and derives some error estimates. We have the following simple lemma.

Lemma 1. SVDT is equivalent to projecting the given data onto the space spanned by the 'significant' singular vectors, $\overline{U}$ in U.

Proof of Lemma 1.

Replace b by $\overline{b} = \overline{U}\,\overline{U}^Tb$. Then the solution $\overline{x}$ of $Ax = \overline{b}$ satisfies $\overline{x} = V\,\Sigma^{-1}\,U^T\,\overline{U}$

$\overline{U}^Tb = V\begin{pmatrix} \overline{\Sigma}^{-1} \\ 0 \end{pmatrix}\overline{U}^Tb = V\begin{pmatrix} \overline{\Sigma}^{-1} & 0 \end{pmatrix}U^Tb = x_{SVDT}$, the

solution of (9).

Q.E.D.

At this point it is appropriate to note that if our problem $Ax = b$ is a discretization of (1) then, Theorem 1 tells us that the requirement that the discreted data,

$$g_n = \sum_{1}^{n} (g_n^T u_j)\, u_j \tag{10}$$

have the property that $|g_n^T u_j| << \sigma_j$ for j near n, is not totally unreasonable if n is 'large'. Theorem 1 requires that

$$\sum_{1}^{\infty} \frac{(g^T u_j)^2}{\sigma_j^2} < \infty \tag{11}$$

Obviously the sum in (11) can be finite only if the individual terms in the sum go to zero. (Note that in (10) we have used u_j to denote a vector, and in (11) to denote a function.) Thus, chopping makes sense in this setting, although for a general problem (2) there is no a priori reason why the above requirement is reasonable. Varah [8] estimated the effect of roundoff error on the SVDT as M/σ_k where M depends on the machine arithmetic used and in (9) σ_j has been set to zero for $j \geq k$. He did not discuss the effect of noise in the data, although roundoff is frequently given a uniformly distributed random model, Voevodin [20].

Suppose we are given $b_\epsilon = b + \epsilon$ where ϵ is randomly generated noise, with mean 0. For large n we therefore expect ϵ to have componentwise alternating signs. What is the effect of ϵ? What does a SVDT do to such an error? Since our system is linear, the additional error in the solution of (9) due to ϵ is

$$\frac{\epsilon_x}{||x||} = V \begin{pmatrix} \Sigma^{-1} \\ 0 \end{pmatrix} \frac{U^T \epsilon}{||x||} \tag{12}$$

Is (12) small? Clearly, (12) will be zero only if ϵ is orthogonal to the vectors u_j, $1 \leq j \leq k$. For (12) to be small we must have ϵ essentially orthogonal to these u_j. Any nonzero projection of ϵ onto these u_j must be sufficiently small with respect to the corresponding projec-

tion of the true data, b, onto these vectors. If we recall our 2 premises, (1) our solution is smooth and (2) the error is highly oscillatory, then this requirement that (12) be small translates into a requirement that the u_j, $1 \leq j \leq k-1$ do not pick up highly oscillatory behavior.

Consider for example the oscillatory matrices Gantmacher [21] or at least matrices with the following oscillation property.

Definition 1. A matrix A with eigenvalues λ_j + j = 1,...,n has the oscillation property if the following is true. There are $m \leq n$ distinct eigenvalues μ_j + j = 1,...,m, and there is a corresponding complete basis of eigenvectors such that each of the basis vectors corresponding to μ_j has exactly (j-1) variations of sign in its coordinates.

Definition 2. A *rectangular matrix* A *is called* totally nonnegative (totally positive) if all its minors of any order are nonnegative (positive).

Definition 3. A matrix A is called oscillatory if A is totally nonnegative, A is nonsingular, and all the elements in the principal diagonal and the first superdiagonal and subdiagonal are nonzero.

Oscillatory matrices arise in the study of small vibrations of elastic systems Gantmacher and Krein [22]. Oscillatory matrices have the oscillation property.

Theorem 3 Gantmacher and Krein [22]. An oscillatory matrix A always has the oscillation property.

Observe that if the matrix A in (2) has the oscillation property, then SVDT can remove the effects of highly oscillatory noise whenever the desired solution is smooth. However, if A does not have this property SVDT may have little effect on the noise. Before presenting an example to demonstrate this we note the following. In Varah [8], he considered 3 classical

continuous, ill-posed problems (1) harmonic continuation, (2) the inversion of a LaPlace transform, and (3) the backwards heat equation. It is interesting to note that although he did not explicitly mention measurement errors, the discrete versions of these 3 problems that he used in his computations each have the oscillation property stated in Definition 1. In fact the matrices used for the LaPlace transform and for the heat equation are both totally positive and hence oscillatory. Thus, each problem in Varah [8] was expanded in a frequency or oscillation oriented basis, and truncating the singular value decomposition was equivalent to removing the high frequency components from the data. In such a situation SVDT achieves our stated objectives.

Now, however, consider the following discussion that shows, SVDT, may not mollify the effects of noise in the data. We consider the family of circulant matrices Gray [23].

Definition 4. An $n \times n$ matrix A is a right circulant if and only if $a_{i-h,j-h} = a_{i,j}$ for all integers $0 \leq i, j, h \leq n-1$ where the indices are computed modulo n.

For example the 4×4 matrix

$$C = \begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix}$$

is a right circulant. A circulant has only n independent entries. Each successive row is the successive cyclic permutation of the elements in the first row of the matrix. The sum, product, transpose and inverse of a circulant matrix is a circulant matrix. Observe that circulants occur in the problem of analytic continuation Varah [8]. Given a harmonic function $u(r,\theta)$ in the unit circle with known values for some $r < 1$, $g(\theta) = u(r,\theta)$, find its values $f(\theta)$ on the unit circle, $r = 1$. f and g satisfy,

$$g(\theta) = \frac{1}{2\pi}\int_0^{2\pi} \left[\frac{1-r^2}{1 - 2r\cos(\theta-\theta_1) + r^2}\right] f(\theta_1)\, d\theta_1 \qquad (13)$$

If we discretize at equally spaced points using the trapezoidal rule and the fact that the kernel and the solution are periodic, we obtain a symmetric circulant equation as an approximation to (13).

The fact that the right circulants form a commutative ring can be verified directly. Let C be a right circulant whose first row is $(c_0, c_1, \ldots, c_{n-1})$. Let $r_j = \exp(2\pi ij/n)$ where $i = \sqrt{-1}$, $0 \leq j \leq n - 1$. The r_j are the n, nth roots of unity. Then the vectors $v_j = w_j/\sqrt{n}$ where

$$w_j^T = (1, r_j, r_j^2, \ldots, r_j^{n-1}) \ , \ 0 \leq j \leq n - 1 \tag{14}$$

form a unitary, eigenvector basis for C. The corresponding eigenvalues are

$$\lambda_j = \sum_{k=0}^{n} c_k r_j^k \ . \tag{15}$$

If C is symmetric, then we must have

$$c_k = c_{n-k} \tag{16}$$

for $k = 1, \ldots, (n/2)-1$ if n is even, and for $k = 1, \ldots, [n/2]$ if n is odd. Note that c_0 and, if n is even, $c_{n/2}$, are singletons. In the symmetric case we therefore get

$$\lambda_k = \lambda_{n-k} \tag{17}$$

for $k = 1, \ldots, (n/2)-1$, if n is even, and for $k = 1, \ldots, [n/2]$ if n is odd. Note that λ_0 and, if n is even, $\lambda_{n/2}$, have multiplicity 1.

We want to construct an ill-posed symmetric circulant matrix that does not have the oscillation property in Definition 1. Clearly the vectors w_j as a function of increasing j have the appropriate oscillatory behavior; the $|\lambda_j|$ however do not have to decrease in size as $j \uparrow$. To verify this, consider the following lemma.

Lemma 2. Let n be odd. Given any sequence of [n/2] + 1 real numbers $\lambda_0,\ldots, \lambda_{[n/2]}$, there is a real, symmetric circulant with eigenvalues $\lambda_0,\ldots, \lambda_{[n/2]}$, and $\lambda_{n-j} = \lambda_j$ for j = 1,..., [n/2].

Proof: To verify Lemma 2 we first note from (15) - (17) that the λ_j would have to satisfy with q = [n/2]

$$\lambda_o = c_0 + 2 \sum_{1}^{q} c_k$$

$$\lambda_j = c_0 + 2 \sum_{1}^{q} c_k \cos\frac{2\pi kj}{n}, \quad 1 \le j \le q \tag{19}$$

Observe that the jth row of the coefficient matrix of (19), if we let our unknown vector be $(c_0/2, c_1, \ldots, c_q)$, is the function cos $2\pi jt$ evaluated at equally spaced points in the interval [0,1]. But we know, see for example Bloomfield [24] these functions are orthogonal over equally spaced points. Therefore, (19) has a unique solution $c_0,\ldots,c_q$ for any set of λ_j. Now define C to be the circulant whose first row is $c_0, c_1, \ldots, c_q, c_q, c_{q-1},\ldots, c_1$.

Q.E.D.

Example 1.

By Lemma 2, for any odd n there is a symmetric circulant C_ϵ with eigenvalues $\lambda_0 = \epsilon$, $\lambda_{[n/2]} = \lambda_{[n/2]+1} = 1$, and the other $\lambda_j > \frac{1}{2}$ as specified. For the SVD of C_ϵ we obtain $\sigma_1 = \sigma_2 = 1$, $\sigma_k = \lambda_{j(k)}$, $\sigma_k \downarrow$, k = 2,..., n - 1, and $\sigma_n = \epsilon$. Using SVDT, and setting $\sigma_n = 0$, we would take out of the basis the vector $w_1 = (1,1, .., 1)/\sqrt{n}$. Obviously this would have little effect on the error since it would not affect any of the oscillatory basis vectors w_j, j > 1.

Thus, an ordering of the $\lambda_j(C)$ according to decreasing magnitude, does not necessarily correspond to increasingly oscillatory patterns in the eigenvectors. There is, in general no direct link between a frequency space analysis of a given problem and its eigensystem. The

circulants give us a concrete family in which we can readily see this behavior. We could of course, construct an example trivially. For if A is any oscillation matrix, then A^{-1} is not. However, we introduced the circulants because they give us an interesting class of matrices that we will be able to use again in the section on regularization. Thus, we see that without an oscillation property requirement there is no correlation between the SVDT and the removal of highly oscillatory noise. Since roundoff error also has a random element, this remark also applies to that kind of noise. SVDT yields a numerically well-conditioned problem and this is an important consideration. We can also argue that our solution should not have any significant part on the ill-conditioned subspace that has been projected out. However, in general (for example, for systems without the oscillation property) if we use SVDT the error in the data is not mollified and thus appears in our solution. We would like to have a procedure that would be numerically stable and that also would remove or mollify the error.

3. A Hankel Transformation Procedure. Ekstrom and Rhoads [9].

First we restate our assumptions and our objective. We are assuming the desired solution is smooth. We are also assuming that our data g (or b) is contaminated by a highly oscillatory, but small amplitude error. We want to replace our original ill-posed problem (1) or (2) by a better conditioned problem that mollifies the effect of the noise in our data on the solution, and that gives us a physically meaningful solution that approximates the true solution in some reasonable sense. As was shown in section 2, the SVDT approach is equivalent to projecting the data onto the 'significant' part of the space. The difficulty as demonstrated in section 2 is that this projection may not mollify the effects of the error, unless the matrix in equation (2) has the oscillation property. The Ekstrom and Rhoads approach [9] also uses a singular value decomposition, but with a weighting system instead of a truncation. This weighting can effectively mollify the effect of the noise. However, by example, we will demonstrate that it can also mollify the desired solution, and this we do not want to do.

The discussion in Ekstrom and Rhoads [9] is restricted to convolution equations. That is in (1)

$$K(t,s) = K(t-s) \tag{20}$$

and in (2)

$$A_{ij} = A_{i-j} \, . \tag{21}$$

However, we will argue that the ideas in [9] can be extended to general matrices, if we use the singular value decomposition. We should note that many interesting applications, for example, numerical differentiation, Cullum [25], Anderssen and Bloomfield [12], and the signal processing of radar, Preiss [26], satisfy (20) or (21). Therefore, even if we restrict the discussion to convolution equations, we are considering a large class of interesting problems. A matrix satisfying (21) is called a Toeplitz matrix. If (20) is satisfied and we discretize (1) appropriately (2) will be a Toeplitz system. However, one should note that preservation of the convolution character depends upon the quadrature formula used.

Definition 5. A matrix A is persymmetric if it is symmetric with respect to the secondary diagonal, $a_{j(n-j+1)}$, $1 \le j \le n$. Clearly, a Toeplitz matrix is persymmetric.

Lemma 3. If $Ax = b$ is a persymmetric matrix and P is the matrix of all zeros except for 1's on the secondary diagonal, then the matrix $H = AP$ is a symmetric matrix. Furthermore, $x = Py$, where y is the solution of $Hy = b$, is the solution of $Ax = b$.

Proof. Since A is persymmetric,

$$a_{ij} = a_{n-j+1,n-i+1} \quad \text{for } 1 \le i, j \le n \, . \tag{22}$$

Moreover, since $H = AP$, $h_{ij} = a_{i(n-j+1)}$. But, (22) implies that $a_{i(n-j+1)} = a_{j(n-i+1)}$, the i,j entry in H^T. Therefore, H is symmetric. Furthermore, if we set $x = Py$, then $Ax = APy = Hy = b$.

Q.E.D.

Therefore, we can transform our system $Ax = b$ into an equivalent symmetric system $Hy = b$. Since A is Toeplitz, H is a Hankel matrix.

Definition 6. A matrix A is an $n \times n$ Hankel matrix if and only if $a_{ij} = h_{i+j}$, $0 \leq i, j \leq n-1$.

In a Hankel matrix the elements along the secondary diagonal are all equal, as well as all elements on diagonals parallel to the secondary diagonal.

Ekstrom and Rhoads [9] make this transformation and work with H. Since H is symmetric it has a full orthonormal eigensystem v_j, $1 \leq j \leq n$. They assume all the eigenvalues λ_j, $1 \leq j \leq n$ are distinct. One can write the solution of $Hy_\epsilon = b + \epsilon$ as

$$y_\epsilon = \sum_1^n \beta_j v_j \tag{23}$$

where

$$\beta_j = \frac{b^T v_j}{\lambda_j} + \frac{\epsilon^T v_j}{\lambda_j} \qquad 1 \leq j \leq n \tag{24}$$

The error in (23) due to ϵ depends upon $\epsilon^T v_j$ and λ_j. To mollify this error, a weighting sequence $W(\lambda_j)$ is introduced. That is, they replace (23) by

$$y(\epsilon, \alpha) = \sum_1^n W_j \beta_j v_j \tag{25}$$

They consider several choices for W, using the formula

$$W_j = \frac{|\lambda_j|}{|\lambda_j| + \alpha_j} \tag{26}$$

where α_j is to be specified. For example, they consider setting α_j equal to some mulitple of the variation in the components of the individual eigenvector, v_j. For example,

$$\alpha_j = \alpha \sum_{i=1}^{n-1} (v_j^{i+1} - v_j^i)^2 \tag{27}$$

or

$$\alpha_j = \alpha \sum_{i=2}^{n-1} (v_j^{i+1} - 2\, v_j^i + v_j^{i-1})^2.$$

A comment on (27) is necessary. These are the relationships stated in Ekstrom and Rhoads. Observe that the first variation since v_j is normalized can be at most 4. Therefore, hidden in the choice of α are the effects of the span of the λ_j and the dimension n, and we cannot say α is always small or large since its size will depend on n and the λ_j scaling. To directly incorporate the effect of n, we write (27) so that the expressions look more like difference quotients, and obtain

$$\alpha_j = \alpha \sum_{i=1}^{n-1} \left[\frac{(v_j^{i+1} - v_j^i)}{1/n} \right]^2 , \tag{28}$$

and similarly for the second variation. Then if $|\lambda_j| < n^2$, (28) can yield a significant damping factor with a small value for α. When we discuss regularization we will be considering derivatives and small values of the regularization parameter, so we choose to use (28) in our discussions to make the relationships between these 2 procedures more transparent. Ekstrom and Rhoads [9] also mention setting $\alpha_j \equiv \alpha$ for all j, this clearly would not connect the weighting to any oscillatory behavior. One must decide how to choose α. Ekstrom and Rhoads give the following heuristic. They observe that in the tests they ran, the residual $||\, b + \epsilon - Ax(\alpha) \,||$ decreased initially as α was reduced and then increased as α was further reduced. Furthermore, the value of α for which the residual was minimized, was a good approximation to the value of α that minimizes the error.

The Ekstrom-Rhoads approach is interesting, especially if one has a medium size problem that is to be solved over and over again. In such a situation the eigenelements can be determined once, stored and used over again on any set of data. For very large problems one would not want to have to compute the eigenelements, but this is also true of a SVDT. However, in many applications the number of data points is reasonable. Use of the filter (28) gives us a way of reducing the effects of the highly oscillatory eigenvectors, and of any small eigenvalues corresponding to smooth eigenvectors. Thus, we can reduce the effect of the measurement errors, and at the same time reduce the numerical instabilities due to the very small eigenvalues. The Ekstrom-Rhoads approach [9] is an SVD with damping of the effects of oscillatory eigenvectors and of the small eigenvalues; in place of the block filter of the SVDT which may have no direct relationship to oscillatory behavior. This damping mimics regularization. Thus, this approach lies somewhere between an SVDT and a regularization. The weighting in Tikhonov regularization, however, has a functional analytic interpretation; whereas it is not clear what a similar interpretation for the weights in (28) would be. Observe in particular that the weights W_j as a function of j need not be a monotone decreasing function of j and in general may have any shape. By example we will show, however, that this approach does not necessarily satisfy our objectives either. We construct a problem where half of the components of 2 of the eigenvectors are oscillatory, but where the other components are not. If such a frequency split occurs, the Ekstrom-Rhoads approach will damp out the projection of the solution on these vectors and may damp out a significant part of the true solution.

Example 2. Let $n = 2m$. Let e be the m - vector of all 1's. Let h be the m - vector (1, -1, 1, -1,...). Then $z_1 = (e, h)$ and $z_2 = (e, -h)$ are n - vectors and the vectors $v_i = z_i/\sqrt{n}$ are orthonormal. There exists a complete orthonormal basis for n - space, $V = \{v_1, v_2, \ldots\}$. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$ be any set of eigenvalues and set $A = V\Lambda V^T$ where Λ is the diagonal matrix whose jth diagonal entry is λ_j.

We want to solve $Ax = b + \epsilon$ where

$b = A(e,0)$ and $\epsilon = (0,\delta h)$. Clearly the solution for $\delta = 0$, equals

$$y = \frac{\sqrt{n}}{2} (v_1 + v_2)$$

and the error in the data,

$$\epsilon = \frac{\sqrt{n}}{2} \delta(v_1 - v_2).$$

Therefore, for example with $\lambda_1 = 1/10$, $\lambda_2 = 1/20$, the computed solution is

$$y_c = \frac{\sqrt{n}}{2} \left[W_1(1 + 10\,\delta)v_1 + W_2(1-20\delta)v_2\right].$$

The corresponding error in the solution is

$$y_c - y = \frac{\sqrt{n}}{2} \left[((W_1-1) + 10\,W_1\delta)v_1 + ((W_2 - 1) - 20W_2\delta)v_2\right]. \qquad (30)$$

The Ekstrom-Rhoads procedure chooses the weights W_i by (28) measuring the oscillation in v_1 and v_2. Using the first variation, we obtain $\alpha_1 = 2\alpha n(n-2)$ and $\alpha_2 = 2\,\alpha\,n^2$. This choice yields damping or small W_i. However, we see in (30) that to get an accurate solution we need each $W_i \sim 1$.

The problem in Example 2 is the mixing of the frequencies. The eigenvectors do not necessarily yield a separation of the frequencies. We note that if A is an oscillation matrix, then the Ekstrom-Rhoads approach is a generalization of SVDT to include the effects of oscillatory behavior. How much of a generalization it is depends upon the particular problem since for any problem (2) obtained from (1), we expect the $\lambda_j^n \to 0$ as $n \to$.

Observe that unlike the SVDT, this approach cannot be interpreted as a modification of the data; it is instead a modification of the operator.

$$H \rightarrow \tilde{H} = V \tilde{\Sigma} V^T$$

where $\tilde{\Sigma}$ is the diagonal matrix whose jth diagonal entry is $(\lambda_j + \alpha_j)$.

To extend these arguments to systems that are not persymmetric, consider the following. Let $A = U \Sigma V^T$ be a singular value decomposition of any matrix A. Thus, V is a system of eigenvectors for A^TA and U a set for AA^T. Our solution of $Ax = b$ is $x = V\Sigma^{-1} U^Tb$. If we know x is not oscillatory we can damp out those v_j and u_j with highly oscillatory behavior. We take

$$x = \sum \frac{(b^Tu_j)\, W_j\, v_j}{\lambda_j} \tag{31}$$

with a weight that measures any oscillation in u_j or v_j. For example let

$$\alpha_j = \alpha n^2 \sum_{k=0}^{n-1} \left[(u_j^{k+1} - u_j^k)^2 + (v_j^{k+1} - v_j^k)^2 \right]. \tag{32}$$

Using (32) we can apply the Ekstrom-Rhoads approach to any problem.

At the beginning of this paper we claimed that regularization is the natural setting for handling noisy information, because it has a direct frequency interpretation, in the next section we will justify this comment. The justification presented is valid only for convolution equations. The prototype convolution equation, which has numerous applications, is the numerical differentiation of tabulated data, and many of the explanatory comments will refer to this problem.

4. Tikhonov Regularization. [10], [28].

To discuss the regularization of equation (1) or (2), we restrict ourselves to convolution equations $K(t,s) = K(t-s)$ or Toeplitz systems, $A = (a_{ij}) = (a_{i-j})$. Thus, system (1) becomes

$$g(t) = \int_0^1 K(t-s)\, f(s)ds \quad , \quad 0 \le t \le 1 \tag{33}$$

We make this restriction because we want to introduce Fourier transforms. However, Tikhonov regularization is applicable to general equations (1).

The Fourier transform of an $L_1(-\infty,\infty)$ function h is defined as

$$\hat{h}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t}\, h(t)\, dt \tag{34}$$

In practice, as in our equation (1), we often have a finite, not infinite interval of integration, and in addition, in our numerical work we deal with sampled information. We will try to incorporate these effects into our remarks, however, we will not present a totally rigorous analysis.

As stated in the previous section an example of (33) is the computation of the derivative of a function. Given g find f where

$$g(t) = \int_0^1 h(t-s)\, f(s)\, ds \tag{35}$$

and $h(t-s) = 0$ for $t < s$, and 1 for $t > s$.

We denote any equation of type (1) by $Kf = g$. The regularization procedure replaces (1) by a one-parameter family of minimization problems which we denote by $P(\alpha)$, $\alpha \geq 0$. For a given $\alpha \geq 0$, we define $P(\alpha)$ as

$$\underset{f}{\text{minimize}} \left[\| Kf - g \|^2 + \alpha \, \Omega(f) \right] \tag{36}$$

where $\| \cdot \|^2$ is some norm measure of the error $(Kf - g)$ in f being a solution of (1), and $\Omega(\cdot)$ is some norm, usually different from $\| \cdot \|^2$, which controls the smoothness of f. For example,

$$\Omega(f) = \int_0^1 f^2 + \int_0^1 (f^{(1)})^2 \tag{37}$$

where $f^{(1)}$ denotes the derivative of f.

Definition 6. The family, $P(\alpha)$, $\alpha \geq 0$ is a regularizing family for $Kf = g$ if (1) as $\alpha \to 0$, the solution f_α of $P(\alpha)$ converges in some reasonable sense to the solution of $Kf = g$, and (2) each problem $P(\alpha)$ is well-posed.

Definition 7. A problem is well-posed if it has a unique solution and this solution depends continuously on the data.

Definition 7 is the Hadamard defintion of well-posedness. However, we should note that a problem can be well-posed but ill-conditioned. For example every system (2) is well-posed, since we assumed A is invertible. However, depending upon the span of the singular values, this system can be arbitrarily ill-conditioned. In numerical work we want well-conditioned, not just well-posed problems and as we look at regularization we will try to see if the approximate problems it generates are in fact well-conditioned.

In practice our problems contain measurement errors. When there is error, we cannot let the regularization parameter α become too small. Note that even without such measurement

errors, there are errors due to the quadrature approximations and due to roundoff. These errors also limit the size of α. In practice we find that the error $||f_{True} - f_\alpha||^2$ reduces initially as α is reduced, reaches a minimum which depends on the problem and the type of error present and then increases again as we continue to reduce α. Procedures for determining an appropriate value of α are given in Wahba [27], Turchin, Kozlov and Malkevich [28], Gordonova and Morozov [29].

One of our stated objectives is to replace our original problem by a better conditioned one. We must determine the 'condition of $P(\alpha)$'. The Euler necessary conditions of optimality for $P(\alpha)$ are

$$(K^*K + \alpha B)\, f = K^*g \quad \text{and } f^{(1)}(0) = f^{(1)}(1) = 0 \tag{38}$$

where B corresponds to Ω. Equation (38) is simply the statement that at an optimum point f of $P(\alpha)$, the first variation of (36) must vanish. If we have equations (35), and (37) the problem of differentiation, equation (38) is the integro-differential equation

$$\int_t^1 \int_0^s f(u)\, du + \alpha f - \alpha f^{(2)} = \int_t^1 g(u)\, du \;,\; f^{(1)}(0) = f^{(1)}(1) = 0 \tag{39}$$

In Cullum [25] an additional term $\int_0^1 f$ was added to the left-hand side of (39). This corresponds to adding $\left(\int_0^1 f\right)^2$ into the expression in (36). To solve (36), we solve (39). Using a quadrature rule and discretizing the derivative in (39), we obtain an algebraic system

$$(H + \alpha B)\, x = b \tag{40}$$

How does the addition of the operator αB affect the condition of K^*K or H? The existing regularization results Anderssen and Bloomfield [30], Hilgers [31] indicate that in many cases the best choice of α is a small number, for example 10^{-8} to 10^{-12}. How can such a small number have a significant effect on the condition of our problem $K^*Kf = K^*g$?

Clearly, if $\Omega(f) = \int_0^1 f^2$ so that $B = I$, the smallest eigenvalue is α and the condition of (38) is essentially $1/\alpha$. So, if α is very small, we have not done much to stabilize our numerical computations. However, if we take $Bf = f - f^{(2)}$, then we can argue that we may have significantly affected the condition. As is also noted in Lee et al [11], where a very different type of regularization analysis is presented, it is important that the operator B in (38) be an unbounded operator whose eigenvalues have a limit point at infinity. The ill-posedness of (1) is due to the compactness of the operator K; compactness implies that the singular values of K have a limit point at 0. We must add on an operator that can counteract this decay. To achieve some feeling for condition, let us consider the circulant matrices again. Note that any circulant is also Toeplitz. If the first row of our circulant C is $c = c_0, \ldots, c_{n-1}$, then we recall the following interesting fact that the jth component of the finite Fourier transform of c, $0 \le j \le n-1$,

$$J_j(c) = \sum_{k=0}^{n-1} \exp\left(\frac{2\pi i k j}{n}\right) c_k = w_j^T c = \lambda_j(c) \quad . \tag{41}$$

That is the Fourier coefficients are just the eigenvalues of C. In (41) w_j is the jth eigenvector of C given in (14).

For $B = f - f^{(2)}$ and $f(1) = f(n)$, a discretized version of B is the symmetric circulant

$$B = \frac{1}{h^2}\begin{pmatrix} 2+h^2 & -1 & 0 & 0 & -1 \\ -1 & 2+h^2 & -1 & 0 & 0 \\ 0 & -1 & 2+h^2 & -1 & 0 \\ 0 & 0 & -1 & 2+h^2 & -1 \\ -1 & 0 & 0 & -1 & 2+h^2 \end{pmatrix} \tag{42}$$

where $h = \Delta t_i$ is fixed.

Observe that B is diagonally dominant, therefore all the eigenvalues are positive in fact

$$\lambda_j = \frac{2}{h^2}\,(1 - \cos(2\pi\, j/n)) + 1 \tag{43}$$

If $h = 1/n$, then $\lambda_j = O(n^2)$, and $\lambda_{n-j} = \lambda_j$ $1 \le j \le n/2$ (if n is even).

Before proceeding we state the following theorem which is proved in Cullum [32].

Theorem 4. Let K be a positive definite, symmetric operator, then the family of equations, $\alpha \ge 0$,

$$(K + \alpha B)f = g,\ f^{(1)}(0) = f^{(1)}(1) = 0 \tag{44}$$

where B is obtained from a norm Ω as in (38) is a regularizing family for (1).

Theorem 4 allows us to use (44) in place of (38). Let C be a symmetric, positive definite, but ill-posed circulant obtained from (1) whose first row is $c_o, c_1, \ldots, c_{n-1}$, and consider the equations

$$(C + \alpha B)x = b$$

Since the matrix $D = C + \alpha B$ is symmetric and positive definite, the condition of D is just the ratio $\max \lambda_j(D)/\min\lambda_j(D)$. All 1-circulants have the same family of eigenvectors so we have $\lambda_j(D) = \lambda_j(C) + \alpha\, \lambda_j(B)$. We know as the discretization is refined, that $\min \lambda_j(C) \to 0$ and $\max \lambda_j(B) \to \infty$. The effect of αB can now be explained by giving an example.

Example 3. Let $\lambda_j(C) = 1/j^q$ and $\lambda_j(B) = j^s$ for some $q > 0$, $s > 0$ and $j = 1,\ldots, n/2$. Note that $\lambda_{n-j}(D) = \lambda_j(D)$, since D is symmetric. Then the minimum of $\lambda_j(D)$ for a given value of α is not less than the minimum of

$$\phi(\omega) = \frac{1}{\omega^q} + \alpha\omega^s \qquad \omega > 0 . \tag{45}$$

But, this minimum occurs when $\omega = [q/\alpha s]^{1/q+s}$, and the corresponding minimum value of (45) is of the form

$$a\alpha^{q/q+s}. \tag{46}$$

For example if $s = 2$ which would correspond to (42) and $q = 2$, that is the eigenvalues of C decay like $1/n^2$, then

$$\lambda_{min}(D) \geq 2\sqrt{\alpha} . \tag{47}$$

Moreover, let $\bar{c} = \sum_{0}^{n-1} |c_j|$, then

$$\lambda_{max}(D) \leq 4\alpha n^2 + \bar{c} . \tag{48}$$

Therefore, the condition of D, to be specific take $s = 2$,

$$\frac{\lambda_{max}(D)}{\lambda_{min}(D)} \leq 2\sqrt{\alpha}n^2 + \frac{\bar{c}}{2\sqrt{\alpha}} . \tag{49}$$

If $\alpha = 10^{-10}$ and $n \leq 10^4$, this condition is $\leq 10^5$.

From (46) we can see the effect on conditioning of increasing the order of the smoothing norm, i.e. increasing s, or of increasing the rate of decay of the eigenvalues of the operator C, i.e. increasing q. If we keep the order of B, s, fixed and increase the rate of decay q, we see that the condition of (44) deteriorates towards $1/\alpha$ as $q \to \infty$. Similarly, if we fix the rate of

decay q and increase the degree of smoothing s we improve the conditioning. These comments are for α fixed. However, as we vary s or q we do not expect the changes in α to be enough to compensate for the changes in (46) so that there is a significant net gain in practice.

The problem of analytic continuation given in section 2, see equation (13) is an interesting example of type (44). The appropriate discretization yields a symmetric circulant matrix whose eigenvalues are $\lambda_j = (r^j + r^{n-j})/(1 - r^n)$ $j = 0,..., n/2$. Since, $r < 1$, for large n, λ_j is of the order r^j $j = 0.,,,, n/2$. (Assume n is even). See Varah [8] for details. If we regularize this problem using (42) the eigenvalues of the regularized matrix are for a given $r < 1$ of the order,

$$\lambda_j(0) \sim r^j + \alpha j^2. \tag{50}$$

As before $\min \lambda_j(D)$ is not smaller than the minimum of

$$\phi(\omega) = r^\omega + \alpha(\omega^2 + 1) . \tag{51}$$

ϕ is convex with a unique minimum $\omega(\alpha)$ satisfying

$$e^{-a\omega} = \frac{2\alpha\omega}{a} \tag{52}$$

where $a = - \ln r$. Taking logarithms of both sides and assuming that $a\omega >> \ln \omega$, we obtain the estimate $\omega(\alpha) \sim \ln \alpha \ / \ \ln r$. Using this estimate we estimate the condition of (44) as $\sim(\alpha(\ln \alpha)^2)^{-1}$. For $\alpha = 10^{-10}$, this quantity is like 10^7. Therefore, the exponential decay in C adversely affects the conditioning and as is discussed in Cullum [32] the corresponding goodness of the regularization approximation.

We now consider the more general case of Toeplitz matrices and convolution operators. We note again that starting with (33), depending on the problem and the quadrature rule used,

we may or may not generate a Toeplitz approximation. In particular if K is Toeplitz so is K^*, however, K^*K need not be Toeplitz, so we have to be careful.

As described above, for any circulant matrix the coefficients of the associated finite Fourier transform and the eigenvalues of the matrix are identical, so that we can discuss the condition of the circulant by using the maximum and minimum of the Fourier transform coefficients or of the eigenvalues.

Symmetric Toeplitz matrices and circulant matrices are strongly related Gray [23], and for such matrices we have the following generalization of the relationship between the Fourier transform of the Toeplitz matrix and the eigenvalues of that matirx.

Theorem 5. Widom [33] Let $T_n = \begin{pmatrix} c_0 & c_1 \cdots & c_{n-1} \\ c_1 & c_0 \cdots & c_{n-2} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ c_{n-1} & & c_0 \end{pmatrix}$

be a symmetric Toeplitz matrix. We assume that the c_n are the Fourier coefficients of an L_2 - function ϕ. Let $m = \text{ess. inf } \phi$ and $M = \text{ess. sup. } \phi$ on the real line.

Then all the eigenvalues of T_n satisfy

$$m \leq \lambda_i(T_n) \leq M \quad . \tag{53}$$

This subject matter is discussed in detail in Grenander and Szego [34]. In our case $M = +\infty$ so we also need the following theorem.

Theorem 6 Grenander and Szego [34]. If A is any Hermitian matrix for which the sum of the moduli of the elements in each row have a common bound,

$$\sum_{j=1}^{n} |a_{ij}| \leq M \qquad 1 \leq i \leq n,$$

Then

$$|\lambda_j(A)| \leq M \qquad 1 \leq j \leq n.$$

Theorem 6 is simply a statement about matrix norms, see Wilkinson [35], for example.

It should be emphasized that Theorem 5 is not directly applicable to our problem because it applies to finite Toeplitz matrices that are sections of infinite matrices. As we increase n in our discrete approximations to (33) we are not simply constructing higher dimensional sections of one infinite dimensional Toeplitz matrix, so the following discussion is somewhat heuristic. Thus, we will consider the Fourier transform of our operator $K^*K + \alpha B$ and replace the normal definition of conditioning of a system of linear equations by the ratio of the maximum of the Fourier transform of our operator to its minimum. This is sometimes called the dynamic range, Makhoul [36].

We want to solve (33); we assume K is symmetric and positive definite. The backward heat equation is an example of such a problem, Varah [8]. We regularize (33) using Theorem 4 and take the Fourier transform of equation (44), obtaining

$$(\hat{K} + \alpha \hat{B})\hat{f} = \hat{g} \tag{54}$$

where Λ denotes Fourier transform. We have neglected the effect of the finite interval. We are in fact acting as though our equation were

$$\int_{-\infty}^{\infty} K(t-s)\, f(s)\, ds = g(t) \quad . \tag{55}$$

If we apply Theorem 5 to our problem (heuristically) having first used the simplest quadrature rule, namely rectangles, then the Fourier transform of the operator K corresponds to the function ϕ. Thus, we argue that the maximum and minimum of this Fourier transform

of K give us an estimate of the condition of the discrete Toeplitz systems (2). For example for numerical differentiation taking Fourier transforms of (38) we get (54) equals

$$\left[\frac{1}{\omega^2} + \alpha(\omega^2 + 1)\right]\hat{f} = \frac{1}{i\omega}\hat{g} \ .$$

As obtained in (46), the minimum of this transform is $2\sqrt{\alpha}$. Clearly, it has no upper bound. However, in practice the number of data points n is bounded for example $n \leq 10^4$ and we can use Theorem 6 to get an upper bound on the eigenvalues. Let k be an upper bound on K then

$$M \leq \alpha n^2 \, 8 + k \tag{57}$$

Therefore, the condition number of the matrix D generated from (35) satisfies

$$\frac{\lambda_{max}(D)}{\lambda_{min}(D)} \leq 8\sqrt{\alpha}\, n^2 + \frac{k}{\sqrt{\alpha}} \tag{58}$$

If $\alpha = 10^{-8}$ and $n = 10^2$ then this becomes

$$\frac{\lambda_{max}(D)}{\lambda_{min}(D)} \leq 8 + 10^4 k \tag{59}$$

A comment about the preceeding arguments is necessary. In practice see Cullum [25] Hilgers [31], it was proposed that one invert the B operator and obtain a pure integral equation. Cullum [25] performed a nonsymmetric inversion, Hilgers [31] a symmetric one. This eliminates the need to approximate derivatives of f in Bf by differences. This conversion, however, yields an equation whose condition behaves like $1/\alpha$, and we are back to the same effect on the conditioning as using $B = I$, except that the conversion has simultaneously smoothed the data and the equations. In many cases this smoothing effectively removes the noise, and the smoothing has been done in a natural way. Lee et al [11] and Hilgers [31]

mention the question of whether to discretize (38) directly or to invert B first. However, to date, there is no published analysis of this question.

Now let us use Fourier transforms to examine what regularization is doing to any random noise in the data. Anderssen and Bloomfield [12] did this for numerical differentiation, using the 2nd derivative of equation (38). To solve (38), again ignore the fact that we are really on a finite interval, take Fourier transforms and obtain (54).

Then

$$\hat{f} = \frac{\hat{g}}{\hat{K} + \alpha \hat{B}} \tag{60}$$

We can deconvolve (60) only if the right-hand side is in L_1. For the numerical differentiation formula, K is not symmetric and we have to use equation (38) which becomes equation (56)

$$\left[\frac{1}{\omega^2} + \alpha(\omega^2 + 1)\right]\hat{f} = \frac{\hat{g}}{i\omega} \quad .$$

Note that in Cullum [25], $\left[\int_0^1 f\right]^2$ was added to the $P(\alpha)$ minimization, one can show that this improves the conditioning, but it cannot be incorporated into the Fourier transform framework. In this case

$$\hat{f} = \frac{-i\omega \hat{g}}{1 + \alpha(\omega^2 + \omega^4)} \tag{61}$$

For small α and ω, $\hat{f}$ in (61) behaves like $\omega\hat{g}$, for ω large like $\hat{g}/\alpha\omega^3$, with a smooth transition in between. Therefore, $\hat{f} \in L_1$ if $\hat{g} \in L_1$, and deconvolution is legitimate. In particular if $\hat{g}_\epsilon = \hat{g} + \hat{\epsilon}$ then, the error in $\hat{f}$ for large ω

$$\hat{f}_\epsilon - \hat{f} = 0\left(\hat{\epsilon}/_{\alpha\omega^3}\right) \quad . \tag{62}$$

Thus, we have an explicit means of decreasing the effects of highly oscillatory error in the data.

If random noise is present in the measurements so that we have $g_\epsilon = g + \epsilon$ instead of g, then g_ϵ may have a spiky pattern and these spikes may appear in g_ϵ as peaks. But the spikes in g_ϵ should by assumption correspond to high frequencies ω and at high frequencies g_ϵ is modulated by ω^3. One could get higher modulation using higher derivatives.

Thus, the purpose of $\alpha\hat{B}$ is to counteract the decay in $\hat{K}^*\hat{K}(\omega)$ and make the deconvolution legitimate. If B is symmetric Toeplitz and given for example by (42), we know its eigenvalues $\to \infty$ as $n \to \infty$, so $\hat{B}(\omega) \to \infty$ as $\omega \to \infty$. Therefore, we have constructed a smooth filter on the transform of the given data that decays monotonically to zero as $\omega \to \infty$. How quickly it decays depends upon α and the order of B. This filter has the functional analytic interpretation given in (36). Regularization is a modification of our given operator, not just a projection or modification of the data. For numerical differentiation, if $g = \Sigma\, a_n \cos n\pi t$, then as Anderssen and Bloomfield [12] showed the nth Fourier coefficient of the solution f,

$$f_n = \left[\frac{n\pi}{1 + \alpha (n\pi)^2 + \alpha(n\pi)^4} \right] g_n \quad . \tag{63}$$

In numerical work we have to use the finite Fourier transform and we have to understand the effects of having only a finite number of data points from a finite interval instead of the continuous values over the whole real line. The finite interval corresponds to multiplying a function defined on the whole interval by a unit pulse; the finite number of samples introduces an effect called aliasing, when the effects at many frequencies get mapped onto the same frequency. See Bloomfield [24] for details. Therefore, we have to assume that we have sampled at a high enough sampling rate to pick up the maximal frequency in the noise. This is typically feasible since most instruments are band-limited so that the noise in the data is also band-limited. Otherwise, a portion of the noise could be mapped back into the lower frequencies which should represent the solution.

Let us summarize the above comments. If the error is directly related to frequency or oscillation, than a natural basis for our space is the Fourier exponential functions since these are parameterized in the frequency. A smooth weighting of the coefficients in the expansion of given functions in this basis corresponds to damping any highly oscillatory error and not to damping the smooth solution.

Regularization, however, has its own problems, of a different nature than the SVDT. One of these is discussed in Cullum [32] where we consider the choice of B and argue that this choice is important because it affects the condition and the rate of convergence of the solution of the $P(\alpha)$ problem to the solution of the original problem. If there is noise in our data, we know we cannot make α arbitrarily small. Franklin [16] discusses rates of convergence but doesn't explain what is causing this bad behavior. One of the examples is the backward heat equation which Varah [8] solved using the SVDT. We noted earlier that for this problem the eigenfunctions have the oscillation property so there is a direct connection between truncating the eigenvector expansion and in mollifying the noise. Therefore, we would expect SVDT to work better on this problem than regularization.

5. Summary.

Three procedures SVDT, Hanson [7], the Ekstrom-Rhoads, Hankel matrix procedure [9], and Tikhonov regularization [10] have been described and their treatment of highly oscillatory noise in the data described. We have seen that SVDT uses the singular vectors of the operator and is equivalent to projecting the given data onto a subspace of the singular vectors.

It is numerically stable, but requires the amplified projections $\frac{g^T u_j}{\sigma_j}$ to decay

to zero rapidly, and except when the operator involved has the oscillation property gives no explicit means of mollifying the effects of the errors in the data. This latter property was demonstrated by an example.

We have also seen that the Ekstrom and Rhoads procedure [9] which they discuss only for convolution equations, but which could be extended to general systems using the SVD and a weighting which measures variation in either the eigenvectors of K^*K or the eigenvectors of KK^* , is also a SVD approach. The original system is replaced by an equivalent symmetric system and an expansion made in the eigenvectors of this system. The weighting they introduce will damp components corresponding to highly oscillatory eigenvectors and damp the effects of small eigenvalues thereby making the procedure numerically stable. However, as we demonstrated by example the damping proposed may also damp the desired solution due to the mixing of frequencies which can occur when one uses an eigenvector basis. Moreover, the weightings that they introduce do not seem to have functional analytic interpretations.

The Tikhonov regularization procedure was also discussed for convolution equations, and Fourier transforms were introduced. We used the Fourier transform to estimate the condition of the regularization approximations. Using numerical differentiation as an example we showed (as in Bloomfield and Anderssen [12]) that there exists a direct relationship between the damping associated with the regularization parameter α and the elimination of highly oscillatory noise. We also noted, however, that regularization is not necessarily always the method to use, its usefulness depends upon the rate of convergence of the regularization approximations. This is discussed in more detail in Cullum [32].

References

1. D. K. Faddeev and V. N. Faddeeva, (1974) On the solution of linear algebraic systems, USSR J. of Computational Math and Math. Phys. 14 (3), 1-19.

2. E. L. Zhukovskii and V. A. Morozov, (1972), On successive bayes regularization of systems of algebraic equations, USSR J. of Computational Math and Math Physic 12 (2), 222-223.

3. A. N. Tikhonov, (1965) The stability of algorithms for the solution of degenerate systems of linear algebraic equations, J. of Computattional Math and Math. Physics, 5 (4), 181-188.

4. R. Bachmann, F. J. DiSalvo, T. H. Geballe, R. L. Greene, R. E. Howard, C. N. King, H. C. Kirsch, K. N. Lee, R. E. Schwall, H. U. Thomas and R. B. Zubeck, (1972) heat capacity measurements on smaller samples at low temperatures, The Review of Scient. Instruments 43 (2), 205-213.

5. T. C. Huang and William Parrish, (1975) Accurate and rapid reduction of experimental x-ray data, Appl. Phys. Letters, 27 (3), 123-124.

6. G. Backus and F. Gilbert (1970) Uniqueness in the inversion of inaccuate gross earth data, Philosophical Trans. of the Royal Soc., London A266, 123-192.

7. Richard J. Hanson (1971), A numerical method for solving Fredholm integral equations of the first kind using singular values, SIAM J. Numer. Anal. 8 (3), 616-622.

8. J. M. Varah, (1973) On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems, SIAM J. Numer Anal. 10 (2), 257-267.

9. Ekstrom, M. P. and Rhoads R. L., (1974) On the application of eigenvector expansions to numerical deconvolution, J. of Comp. Phys. 14, 319-340.

10. A. N. Tikhonov, (1963), Solution of incorrectly formulated problems and the regularization method, Soviet Math. Dokl. 4, 1035-1038.

11. D. A. Lee, R. M. Potter, W. Perry and W. Schmaedeke, (1975) Some practical aspects of the treatment of ill-posed problems by regularization, ARL 75-0022, Aerospace Research Labs, Wright-Patterson Air Force Base, Ohio 45433.

12. R. S. Anderssen and P. Bloomfield, (1974), Numerical differentiation procedures for non-exact data, Numer. Math 22, 157-182.

13. J. M. Varah (1976), A practical examination of some numerical methods for linear discrete ill-posed problems, T.R. 76-08, Computer Science, The University of British Columbia.

14. A. D. Wilson, T. H. P. Chang, and A. Kern (1975), Experimental scanning electron-beam automatic registration system, J. Va. Sci. Technol. 12 (6), 1240-1245.

15. Chuan C. Chang (1971), Auger electron spectroscopy, Surface Science 25, 53-79.

16. Joel N. Franklin (1974) On Tikhonov's method for ill-posed problems, Math. of Comp. 28 (128), 889-907.

17. Smithies, F. (1958) Integral Equations, Cambridge University Press, London.

18. Charles F. Van Loan, (1976), Generalizing the singular value decomposition, SIAM J. Numer. Anal, 13 (1), 76-83.

19. Charles L. Lawson and Richard J. Hanson (1974), Solving Least Squares Problems Prentice-Hall, Englewood Cliffs, N. J.

20. V. V. Voevodin (1966) The asymptotic distribution of rounding off errors in linear transformations, USSR J. of Computational Math and Math. Physics. 7 (5), 3-17.

21. F. R. Gantmacher (1964), The Theory of Matrices, Vol. II Chelsea Publ., New York.

22. F. R. Gantmacher and M. G. Krein (1963) Oscillating Matrices, Akademie-Verlag, Berlin.

23. Robert M. Gray (1971), Toeplitz and circulant matrices, a review, TR 6502-1, Information Systems Lab. Stanford Electronics Lab. Stanford U., Stanford, Calif.

24. Peter Bloomfield, (1976) Fourier Analysis of Time Series, An Introduction, John Wiley & Sons, New York.

25. Jane Cullum (1971), Numerical differentiation and regularization, SIAM J. Numer. Anal. 8 (2), 254-265.

26. Douglas H. Preiss (1972), The Toeplitz matrix: its occurrence in antenna problems and a rapid inversion algorithm, IEEE Transactions on antennas and propagation, 204-206.

27. Grace Wabha, (1976), A survey of some smoothing problems and the method of generalized cross-validation for solving them, T. R. 457, Department of Statistics, Univ. of Wisconsin.

28. V. F. Turchin, V. P. Kozlov, and M. S. Malkevich, (1971), The use of mathematical - statistics methods in the solution of incorrectly posed problems, Soviet Physics, 13 (6), 681-840.

29. V. I. Godonova, and V. A. Morozov (1971), Numerical parameter selection algorithms in the regularization method, USSR J. of Computational Math. and Math. Phys. 13 (3), 1-9.

30. R. S. Anderssen and Peter Bloomfield, (1974), A time series approach to numerical differentiation, Technometrics, 16 (1), 69-75.

31. John W. Hilgers, (1976) On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems, SIAM J. Numer. Analy. 13 (2), 172-184.

32. Jane Cullum (1977) The choice of smoothing norm in regularization - a key to effectiveness, IBM RC 6511, IBM Research, Yorktown Heights, N. Y.

33. H. Widom, (1965) Toeplitz Matrices in Studies in Real and Complex Analysis (Math Ass. Amer. Studies in Mathematics) I. I. Hirschmann Jr. Ed. Prentice-Hall, Englewood Cliffs, N. J., 179-209.

34. Ulf Grenander and Gabor Szego, (1958) Toeplitz Forms and Their Applications, The University of California Press, Berkeley, Calif.

35. J. H. Wilkinson (1965), The Algebraic Eigenvalue Problem, Clarendon Press, Oxford.

36. John Makhoul (1975), Linear Prediction: A Tutorial Review, Proc. of the IEEE 63 (4), 561-580.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 77-0880

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

ILL-POSED PROBLEMS, REGULARIZATION, AND SINGULAR VALUE DECOMPOSITIONS

5. TYPE OF REPORT & PERIOD COVERED: Interim

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Jane Cullum

8. CONTRACT OR GRANT NUMBER(s): F44620-76-C-0022

9. PERFORMING ORGANIZATION NAME AND ADDRESS

International Business Machines Corp
T. J. Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A4

11. CONTROLLING OFFICE NAME AND ADDRESS

Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332

12. REPORT DATE: April 1977

13. NUMBER OF PAGES: 37

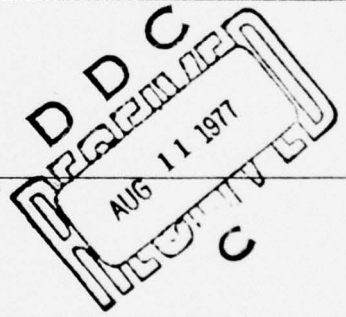
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.



17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

regularization, ill-posed problems, singular value decompositions, numerical differentiation, integral equations of the first kind, deconvolution, convolution equations, smoothing.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Abstract: We consider ill-posed problems of the form $g(t) = \int_0^1 K(t,s)f(s)ds$ and their discrete approximations obtained by quadrature, $Ax = b$. We assume that our desired solution f is smooth and that our data g is measured experimently and contains highly oscillatory noise. With theorems and examples we demonstrate the effect of each of these procedures, the singular value decomposition with truncation, (SVDT) a Hankel transformation with damping, and the Tikhonov regularization procedure, on such noise in the data. We demonstrate that in general, regularization is the most natural setting for mollifying the effects of such noise. However, for certain problems SVDT is equally suitable and in fact may be better if the rate of convergence of the regularization procedure is too slow.

DD 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)